

Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin

André Marques^{a,b}, Tiago Ribeiro^{a,b}, Pavel Neumann^c, Jiří Macas^c, Petr Novák^c, Veit Schubert^b, Marco Pellino^b, Jörg Fuchs^b, Wei Ma^b, Markus Kuhlmann^b, Ronny Brandt^b, André L. L. Vanzela^d, Tomáš Beseda^e, Hana Šimková^e, Andrea Pedrosa-Harand^{a,1}, and Andreas Houben^{b,1}

^aLaboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, 50670-420 Recife, Pernambuco, Brazil; ^bLeibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, 06466 Stadt Seeland, Germany; ^cLaboratory of Molecular Cytogenetics, Institute of Plant Molecular Biology, Biology Centre of the Academy of Sciences of the Czech Republic, 370 05 Ceske Budejovice, Czech Republic; ^dLaboratory of Cytogenetics and Plant Diversity, Department of General Biology, State University of Londrina, 86057-970, Londrina, Paraná, Brazil; and ^eCenter of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, CZ-78371 Olomouc, Czech Republic

Edited by James A. Birchler, University of Missouri-Columbia, Columbia, MO, and approved September 11, 2015 (received for review June 23, 2015)

Holocentric chromosomes lack a primary constriction, in contrast to monocentrics. They form kinetochores distributed along almost the entire poleward surface of the chromatids, to which spindle fibers attach. No centromere-specific DNA sequence has been found for any holocentric organism studied so far. It was proposed that centromeric repeats, typical for many monocentric species, could not occur in holocentrics, most likely because of differences in the centromere organization. Here we show that the holokinetic centromeres of the Cyperaceae *Rhynchospora pubera* are highly enriched by a centromeric histone H3 variant-interacting centromere-specific satellite family designated “Tyba” and by centromeric retrotransposons (i.e., CRRh) occurring as genome-wide interspersed arrays. Centromeric arrays vary in length from 3 to 16 kb and are intermingled with gene-coding sequences and transposable elements. We show that holocentromeres of metaphase chromosomes are composed of multiple centromeric units rather than possessing a diffuse organization, thus favoring the polycentric model. A cell-cycle-dependent shuffling of multiple centromeric units results in the formation of functional (poly)centromeres during mitosis. The genome-wide distribution of centromeric repeat arrays interspersing the euchromatin provides a previously unidentified type of centromeric chromatin organization among eukaryotes. Thus, different types of holocentromeres exist in different species, namely with and without centromeric repetitive sequences.

centromere | satellite DNA | holokinetic | chromosome | evolution

The centromere is the chromosome region where the microtubules attach to the chromatids to enable their movement to the daughter cells during mitosis and meiosis. This region is often enriched in repetitive DNA families, with satellite DNAs (satDNAs) and transposable elements, such as *Ty3/gypsy*-like retrotransposons being the most frequent ones in plants. However, there is little sequence conservation among species (1, 2), and centromere-specific sequences are neither sufficient nor required for the centromere identity (3). Instead, the assembly site for the kinetochore complex of most active centromeres is epigenetically determined by the chromosomal location of the centromeric histone H3 variant CENH3, also known as “CENP-A” (4). Nevertheless, some evolutionary preferences seem to exist, and long-established centromeres are frequently formed on long arrays of satDNAs and/or transposable elements (1).

In certain independent eukaryotic lineages holocentric (also called “holokinetic”) chromosomes occur. These holocentrics lack a primary constriction, and they form kinetochores distributed along almost the entire poleward surface of the chromatids, to which the spindle fibers attach (5, 6). The best-analyzed organisms possessing holocentric chromosomes are the nematode *Caenorhabditis elegans* (7, 8) and some species of the plant genus *Luzula* (9–11). In both

cases, a longitudinal CENH3-positive centromere structure was observed during mitosis. In the rush *Luzula* (Juncaceae), the longitudinal centromere forms a groove (here referred as the “centromere groove”) in each sister chromatid along almost the whole metaphase chromosome except for the most terminal regions (9–11). Recently, a similar centromere organization was found in the sedge species *Rhynchospora pubera* (12). The absence of CENH3 and the centromeric protein C (CENP-C) in some lineages of holocentric insects (13) challenges the general notion of a conserved molecular composition of centromeres in mono- and holokinetic chromosome species. Furthermore, no centromere-specific repeat has been identified thus

Significance

Holocentric chromosomes are characterized by kinetochore activity along each sister chromatid. Although the kinetochore structure seems to be well conserved, as in monocentric organisms, the organization of holocentromeres is still elusive, and no centromeric repeat has been found associated with centromeric histone H3 variant-positive centromeric nucleosomes for any holocentric organism studied hitherto. We demonstrate that holocentrics of the sedge (Cyperaceae) *Rhynchospora pubera* possess different classes of centromere-specific repeats. Holocentromeres are composed of multiple centromeric units interspersing the gene-containing chromatin, and, as a functional adaptation, a cell-cycle-dependent shuffling of centromeric units results in the formation of functional (poly)centromeres during cell division. The genome-wide distribution of centromeric repeat arrays interspersing the euchromatin provides a previously unidentified type of centromere organization.

Author contributions: A.M., A.P.-H. and A.H. designed research; A.M., T.R., P. Neumann, J.M., P. Novák, V.S., M.P., J.F., W.M., M.K., R.B., A.L.L.V., T.B., and H.S. performed research; A.M., P. Neumann, J.M., P. Novák, and V.S. analyzed data; and A.M., P. Neumann, J.M., A.P.-H., and A.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Assembled BAC sequences are available through the iPlant Data Store and can be accessed via iPlant Discovery Environment or at https://de.iplantcollaborative.org/dl/d/8258A143-C5F5-4DF1-84F2-88C948E8E8F/R_pubera_holocentromeres_data.rar. Original ChIP-sequencing sample data, original Illumina sequencing data for the genomic DNA and BACs, and RNA-sequencing data have been deposited at the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) [accession nos. PRJEB9647 (ChIP-sequencing sample data), PRJEB9643 (Illumina sequencing data for genomic DNA), PRJEB9649 (Illumina sequencing data for BACs), and PRJEB9645 (RNA-sequencing data)]. CENH3 cDNA sequences variants were deposited in the GenBank database [accession nos. KR029618 (RpCENH3_1) and KR029619 (RpCENH3_2)].

¹To whom correspondence may be addressed. Email: andrea.pedrosaharand@pesquisador.cnpq.br or houben@ipk-gatersleben.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1512255112/-DCSupplemental.

far for any species possessing holocentric chromosomes (14–17). For instance, Heckmann, et al. (15) have characterized the high-copy fraction of the *Luzula elegans* genome in detail, and none of the identified repeats showed colocalization with the holokinetic centromere. Even in *C. elegans*, where robust studies [Chip-on chip (ChIP-chip) and ChIP sequencing (ChIP-seq)] were performed, no centromere-specific repeats were identified (14, 16).

ChIP-chip experiments suggested that nearly half of the *C. elegans* genome may be associated with CENH3 domains, although only 4% of them form the holocentromeres. Therefore, CENH3 nucleosomes may assemble at random positions (14). However, a recent study based on native CENH3 ChIP sequencing proposed that the holocentromeres in *C. elegans* consist of about 700 individual centromeric units distributed along the length of the chromosomes. Each of these units is formed by only one CENH3-containing nucleosome, where microtubules attach during cell division. These centromeric sites coincide with transcription factor hotspots which are occupied by many transcription factors, without having high binding affinity for any of them (16). The observation that any sequence of *C. elegans* can be propagated as an extra-chromosomal array suggests that no specialized sequences are required for the segregation of holocentric chromosomes (18). On the other hand, that DNA arrays did not segregate with the same fidelity as normal chromosomes indicates that these extrachromosomal arrays lack certain features that promote mitotic stability of wild-type chromosomes (19).

To test the assumption that all holocentromeres are devoid of centromere-specific repeats, we analyzed *R. pubera* (12, 20–22). We report here on the first (to our knowledge) conserved centromere-specific repeats of a holocentric species and propose a model of chromosome organization for species with centromeric repeats distributed throughout the entire genome.

Results

Satellite Repeat Tyba Shows a Holocentromere-Specific Localization.

To identify putative centromeric repeats in the genome of the holocentric plant *R. pubera*, with diploid chromosome number $2n = 10$ and a genome size of the unreplicated reduced chromosome complement ($1C = 1.61$ Gbp), we performed high-throughput shotgun sequencing. A randomly sampled proportion (8.89 million) of generated paired-end reads then was subjected to bioinformatic analysis, implemented within the clustering-based repeat identification pipeline (23, 24). This analysis resulted in thousands of clusters, or groups of reads, with overlapping sequences, each representing a single repeated element or part of it. After repeat classification within major clusters, the global repeat composition of the genome was determined by taking into account the sizes (number of reads) of individual clusters, which are proportional to the genomic abundance of the corresponding repeats.

The *R. pubera* genome was found to be relatively poor in repeated sequences, with highly and moderately repetitive elements represented by clusters with genome proportions of at least 0.01%, collectively making up 41.16% of the genome. The majority of these sequences were classified into the major classes of repetitive DNA. DNA transposons (8.81% of the genome), with more than half represented by miniature inverted-repeat transposable elements (MITEs), accounted for the most frequent repetitive elements, followed by *Ty1/copia* (8.68%) and *Ty3/gypsy* (5.14%) LTR retrotransposons. Other classes of repetitive DNA, such as non-LTR retrotransposons, pararetroviruses, *hAT* DNA transposons, and helitrons were found in much lower genomic proportions (*SI Appendix, Table S1*).

All satDNA reads were grouped into two highly abundant clusters, representing 2.24% and 1.36% of the genome. Because of the high similarity of reads (~70%) present in both clusters, they were considered as two subfamilies of the same satDNA repeat, named “Tyba” (meaning “abundance” in Tupi-Guarani, a language spoken by many Brazilian native tribes), and were designated “Tyba1” and “Tyba2.” The consensus monomer length of Tyba1 and -2 is 172 bp (*SI Appendix, Fig. S2C*), which is in the range of

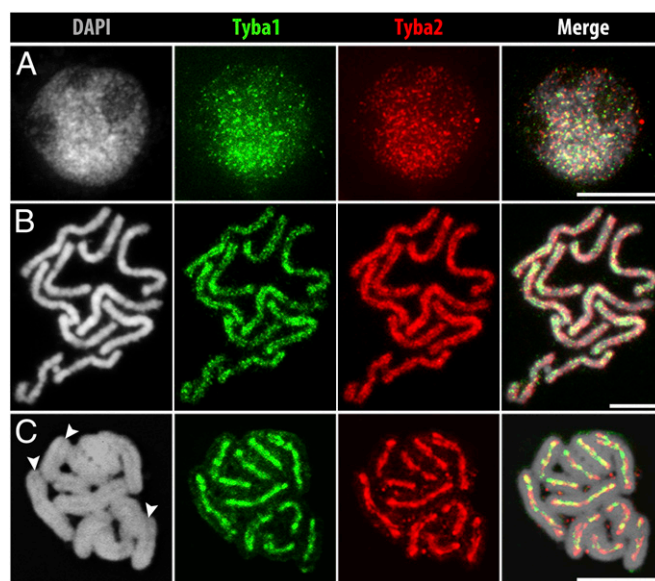


Fig. 1. FISH localization of Tyba1 and -2 in *R. pubera*. (A) Hybridization signals of both Tyba subfamilies in an interphase nucleus show a genome-wide dot-like labeling. (B) Prometaphase chromosomes show a line-like but dispersed labeling on the poleward surface of each chromatid and (C) a distinct labeling along the centromere groove of both sister chromatids during metaphase. Arrowheads in C indicate grooves. (Scale bars: 5 μ m).

monomer lengths of many other known centromeric satDNAs (1, 25, 26). After FISH with Tyba1- and Tyba2-specific probes, interphase nuclei showed dispersed dot-like signals (Fig. 1A), contrasting with the typical strong clustered distribution of satDNA in interphase nuclei of other species. With the onset of chromosome condensation, Tyba signals associate and form dot-like lines along the sister chromatids (Fig. 1B and *SI Appendix Fig. S14*). At metaphase, all chromosomes showed line-like FISH signals colocalizing with the longitudinal DAPI-negative centromere groove (Fig. 1C and *Movie S1*). The hybridization intensity of Tyba1 and -2 probes varied along the chromatids, but only minor labeling was found outside the groove. Both the repetitive nature and the chromosomal distribution of Tyba1 and -2 indicate that this satDNA family is centromere specific in *R. pubera*. In contrast, other high-copy sequences, such as MITE DNA transposons, showed dispersed labeling throughout the chromosomes (*SI Appendix, Fig. S1B*).

Because no other high-copy satellite repeat other than Tyba was found in *R. pubera*, and no heterochromatic domains are visible throughout the cell cycle, we asked whether this genome is organized on a large scale into euchromatin- and heterochromatin-enriched subregions. Therefore, we applied antibodies against typical euchromatin- and heterochromatin-associated histone methylation marks (H3K4me3 and H3K9me2, respectively). In agreement with other holocentric species (15, 27) the euchromatin and heterochromatic domains in *R. pubera* were found dispersed along the entire chromosomes (*SI Appendix, Fig. S1 C and D*). Thus, on a large scale, euchromatin and heterochromatic domains are interspersed.

Tyba satDNA Interacts with CENH3-Containing Nucleosomes.

The centromere specificity of Tyba repeats was tested by colocalization analysis and ChIP with a *R. pubera*-specific CENH3 antibody. First, the pollen mother cell transcriptome of *R. pubera* was determined, and assembled RNA-sequencing (RNA-seq) reads were used to identify *CENH3*. RT-PCR revealed, in addition to the 680-bp-long *CENH3*-like sequence (*RpCENH3_1*, GenBank accession no. KR029618), a second *CENH3*-like variant (*RpCENH3_2*, GenBank accession no. KR029619) of 708 bp, characterized by an insertion in the C-terminal tail after the stop codon (*SI Appendix, Fig. S2A*). Both are expressed mainly

in anthers and roots but less in leaf tissue, most likely because of the higher number of dividing cells in anthers and roots (*SI Appendix, Fig. S2D*). Alignment of the *R. pubera* CENH3 candidates with CENH3s of other plant species supported the correct identification (*SI Appendix, Fig. S2B*). Phylogenetic analysis grouped both RpCENH3s as a sister branch of Juncaceae and other monocot CENH3s (*SI Appendix, Fig. S2E*) (transcriptome assemblies, alignments, and trees are available through the iPlant Data Store and can be accessed via iPlant Discovery Environment or at https://de.iplantcollaborative.org/dl/d/8258A143-C5F5-4DF1-84F2-88C94BE8EA8F/R_pubera_holocentromeres_data.rar).

Next, an antibody (anti-RpCENH3) designed to recognize both CENH3 variants of *R. pubera* was generated and used for immunostaining. A specific labeling of the centromere groove was found at metaphase (*SI Appendix, Fig. S3A* and *Movie S2*). In interphase nuclei, a dispersed distribution of CENH3 foci was found, whereas prophase chromosomes displayed line-like signal arrangements (*SI Appendix, Fig. S3 B and C*). Colocalization experiments using the CENH3-specific antibody and a Tyba-specific FISH probe revealed a coincidence of both signal patterns along the holocentromeres of metaphase chromosomes (*Fig. 2 A and B* and *Movie S3*) and in interphase nuclei (*Fig. 2C*). Detailed analyses using super-resolution microscopy at a lateral resolution of ~ 120 nm quantified an overlap of 57.8% of CENH3 signals with Tyba signals and a 69.1% overlap of Tyba with CENH3 in interphase ($n = 4$). Within the holocentromeres of metaphase chromosomes ($n = 8$) these values amounted to 53.6% and 76.0%, respectively. Double-labeling experiments on extended chromatin fibers confirmed the segmental overlap of the two signal types (*Fig. 2D*). Hence, as in other species e.g., *Arabidopsis*, maize, rice, and humans (28–31), not all centromeric repeat copies interact with CENH3-containing chromatin, and vice versa. Finally, the association of the satDNA with functional centromeres was analyzed by CENH3-ChIP. Quantitative PCR showed a higher than fourfold enrichment of Tyba1 and -2 in the

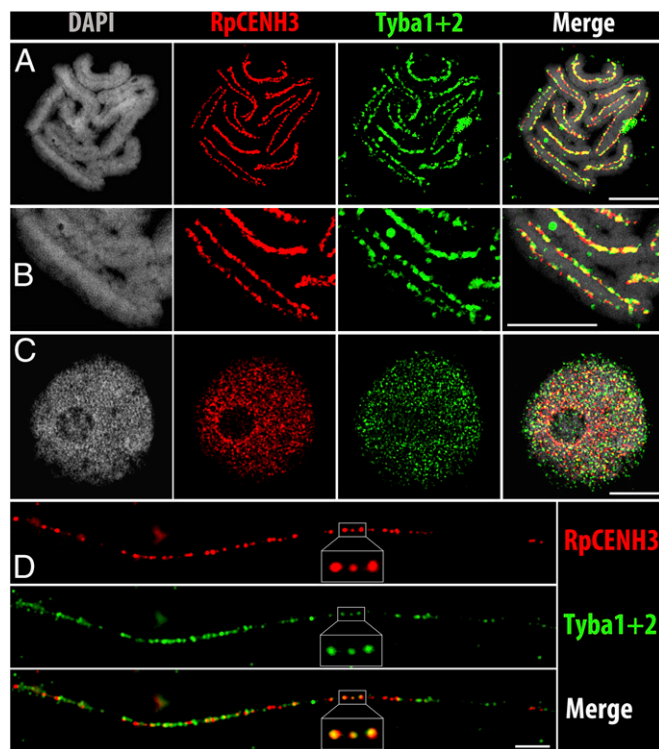


Fig. 2. Colocalization of RpCENH3 and Tyba1 and -2 on *R. pubera* holocentromeres. (A) Metaphase. (B) Enlarged metaphase chromosome. (C) Interphase nucleus. (D) Extended chromatin fiber. A–C are superresolution microscopy (SIM) images. (Scale bars: 5 μ m.)

RpCENH3-ChIP-DNA compared to the input DNA (*SI Appendix, Fig. S2F*). Analysis of ChIP-seq reads demonstrated that Tyba1 and -2 showed the highest level of association in the immunoprecipitated fraction, with 10- and 15-fold enrichment, respectively (*Fig. 3A*). Thus, both members of the Tyba satDNA interact with CENH3-containing nucleosomes.

To validate the tandem-repeat organization and to check whether Tyba is present also in related species, we performed PCR and Southern hybridization using *DraI*-digested genomic DNA of different species of *Rhynchospora* and four other Cyperaceae genera. PCR using Tyba1-specific primers resulted in amplification products in all tested species. However, Southern hybridizations using Tyba1 from *Rhynchospora tenuis* exhibited only tandem repeat-typical ladder-like patterns in all *Rhynchospora* species (*SI Appendix, Fig. S3D*), indicating a conserved abundance of this repeat within this genus. Indeed, immuno-FISH using *R. pubera* Tyba probes and antibodies against CENH3 confirmed a colocalized holocentric signal distribution in *R. tenuis* and *R. ciliata* (*SI Appendix, Fig. S3 E and F*) similar to that in *R. pubera*. Additionally, as reported for other centromeric sequences (32), Tyba1 and -2 are transcriptionally active in all tissues of *R. pubera* analyzed by RT-PCR (*SI Appendix, Fig. S3G*) and RNA-seq search.

Although Tyba is clearly the most abundant centromeric repeat found in *R. pubera*, detailed analysis of ChIP-seq data revealed that CENH3-containing chromatin is also associated with at least three additional families of repetitive sequences: the *Ty3/gypsy* LTR retrotransposon CRRh and two different Tyba-containing repeats, hereafter referred to as “TCR1” and “TCR2” (*Fig. 3A*). The CRRh retrotransposon family, representing about 0.2% of the genome, was found to be heterogeneous, including two autonomous elements (CRRh-1 and CRRh-2) and three nonautonomous elements (noaCRRh-1, noaCRRh-2, and noaCRRh-3). Each autonomous element possessed a single ORF of 1,473 and 1,463 codons, respectively. The putative polyprotein sequences were relatively divergent (59.5% identity), but they both contained all the domains necessary for replication and integration (Gag, protease, reverse transcriptase, RNase H, and integrase) (*SI Appendix, Fig. S4A*). In contrast, the nonautonomous elements did not encode any of the protein domains found in the autonomous elements, but each of them possessed an ORF of unknown function (*SI Appendix, Fig. S4A*). Although the internal sequences of noaCRRh-1 and -2 had no significant similarity to either of the autonomous elements, their LTR sequences shared 73.3 and 72.7% similarity, respectively, with the LTR sequence of CRRh-1, suggesting that they are derivatives of CRRh-1. On the other hand, noaCRRh-3 seemed to have originated from CRRh-2, because these elements shared significant similarities at both LTR termini and in a short region in the 5' UTR. Phylogenetic analysis based on reverse transcriptase domain sequences revealed that the CRRh elements belong to the group B of the centromeric retrotransposon of maize (CRM) clade of chromoviruses (*SI Appendix, Fig. S4B*). Similar to other CRM chromoviruses of this group, both CRRh elements lacked the putative targeting domain at the C terminus of integrase (*SI Appendix, Fig. S4E*) (2). In-depth analysis of ChIP enrichment showed that only CRRh-1, noaCRRh-1, noaCRRh-2, and noaCRRh-3 elements are associated with CENH3-containing chromatin, but CRRh-2 is not (*SI Appendix, Fig. S4C*). Inspection of sequences at insertion sites further revealed that all four ChIP-enriched elements are frequently integrated into Tyba arrays (*SI Appendix, Fig. S4D*). The observed number of CRRh insertions into Tyba was 13- to 21-fold higher than expected (P value $< 2.2 \times 10^{-6}$), implying preferential CRRh insertion into Tyba regions. This observation was confirmed experimentally by FISH, which detected CRRh-specific signals intermingled with Tyba signals within the longitudinal centromere groove (*Fig. 3B*), although the CRRh signals were less abundant. Thus, the holocentromeres of *R. pubera* are enriched in different types of repetitive sequences, with the satDNA Tyba being the main sequence contributing to the formation of active centromeric units, and CRRh showing only moderate enrichment.

Unlike CRRh, whose association with Tyba is clearly a result of its integration preferences, Tyba is a constituent of the repetitive

unit in both TCR1 and TCR2. The sequence of the repetitive element TCR1 could be reconstructed in part, because the ~5-kb-long fragments represent around 0.14% of the genome. One end of the reconstructed consensus sequence likely represents the real boundary of the repetitive unit, because the analysis of this end in reads from different genomic loci revealed a high diversity in the flanking sequences, as is typical for insertion sites. The other end possessed a sequence with high similarity to Tyba. Because our attempts to bridge the Tyba region were not successful, the other boundary of the TCR1 repeat remained unknown. Nevertheless, the presence of putative insertion sites at the Tyba-lacking terminus strongly suggests that TCR1 is a transposable element. In contrast, the repetitive element TCR2 (0.02% of the genome) lacked any sign of insertion sites, and both ends of the in silico reconstructed fragment possessed sequences with high similarity to Tyba. PCR with primers designed from the Tyba-unrelated part of the repeat and directed outwards from the repetitive unit amplified three major fragments about 2.5, 5.0, and 7.5 kb in length, suggesting that TCR2 is a tandem repeat. Sequencing of four randomly selected clones revealed that the monomer of TCR2 is 2,551 bp long and possess nine monomers of Tyba.

Centromeric Sequences Are Composed of High-Order Tyba Tandem Repeats Interspersing the Gene-Containing Chromatin. The higher-order organization of centromeric DNA was analyzed using seven Tyba-containing genomic BACs of *R. pubera*. The identified Tyba arrays were found to be rather small, varying from 3 to 16 kb. As typical for centromeric satDNAs (25, 33), we found the Tyba arrays forming high-order repeat (HOR) structures. Pentamers (830–870 bp) were the most frequent HORs in all BACs analyzed, whereas dimers (~344 bp) were found only occasionally in two BACs (8P1 and 23M1) (*SI Appendix, Table S2*). Tyba arrays occurred as a continuous array in five of the seven BACs; in BACs 17C8 and 23H8 the array was divided into subarrays (Fig. 3C and *SI Appendix, Fig. S3H*). The regions between these subarrays did not show similarity to any known sequence.

A close proximity between Tyba and putative centromeric retroelements was found in three BACs (17C8, 9H8, and 8P1). To confirm these elements as centromeric retroelements, their reverse transcriptase-domain sequences were analyzed phylogenetically. Only the element found in BAC 8P1 grouped within other centromeric retroelements, relatively close to the CRRh clade. The

elements found in BACs 17C8 and 9H8 represent most likely a chromovirus of the Tekay clade and a pararetrovirus, respectively (*SI Appendix, Fig. S3H*). Interestingly, in BACs 17C8 and 9H8 Tyba arrays were found very close to the protein domain regions of these elements, but in BAC 8P1 the CRRh-like element was present ~20 kb upstream of the Tyba array (Fig. 3C and *SI Appendix, Fig. S3H*). Thus, although we frequently detected CRRh elements integrated into Tyba arrays (*SI Appendix, Fig. S4D*), we did not observe this integration in the sequenced BACs.

In addition to a long Tyba array, BAC 23M1 showed a short Tyba-like cluster (~700 bp) inserted into an LTR retrotransposon-like sequence (*SI Appendix, Fig. S3H*). In addition, centromeric sequences were generally found to be flanked by gene-coding sequences and other transposable elements (Fig. 3C and *SI Appendix, Fig. S3H*). A transcriptional activity of these genes is likely, because BLAST comparison of these coding sequences found high similarity between them and the pollen mother cell transcriptome.

Discussion

Origin and Genome-Wide Spreading of a Holocentromeric Satellite. How could a centromeric satellite repeat evolve along a chromosome in thousands of interspersed arrays? The chromosomes of *R. pubera* differ in their organization from all other holocentromeric species studied so far because no other repeat-based centromeres have been reported to date. SatDNA evolves according to the principles of concerted evolution. Within a genome, mutations are homogenized among repeats by the mechanisms of nonreciprocal sequence transfer, such as unequal crossover, gene conversion, rolling circle replication, and transposition-related mechanisms (34). Although the centromere traditionally was treated as a region of suppressed recombination, unequal crossing-over, and gene conversion have been identified as the most widespread mechanism involved in satDNA dynamics (35–37). In addition, segmental duplication has been proposed as a mechanism for massive amplification of satDNA arrays (38, 39).

However, the origin of novel satellite repeats remains elusive. Tandem repeats with homology to parts of retrotransposons have been identified in several plants, e.g., potato (40). Here, we found that the centromeric retrotransposon CRRh is very frequently (>40%) integrated into Tyba satellite arrays, although so far there is no evidence for the presence of Tyba inside CRRh. Thus, CRRh most likely did not contribute to the origin and spreading of Tyba.

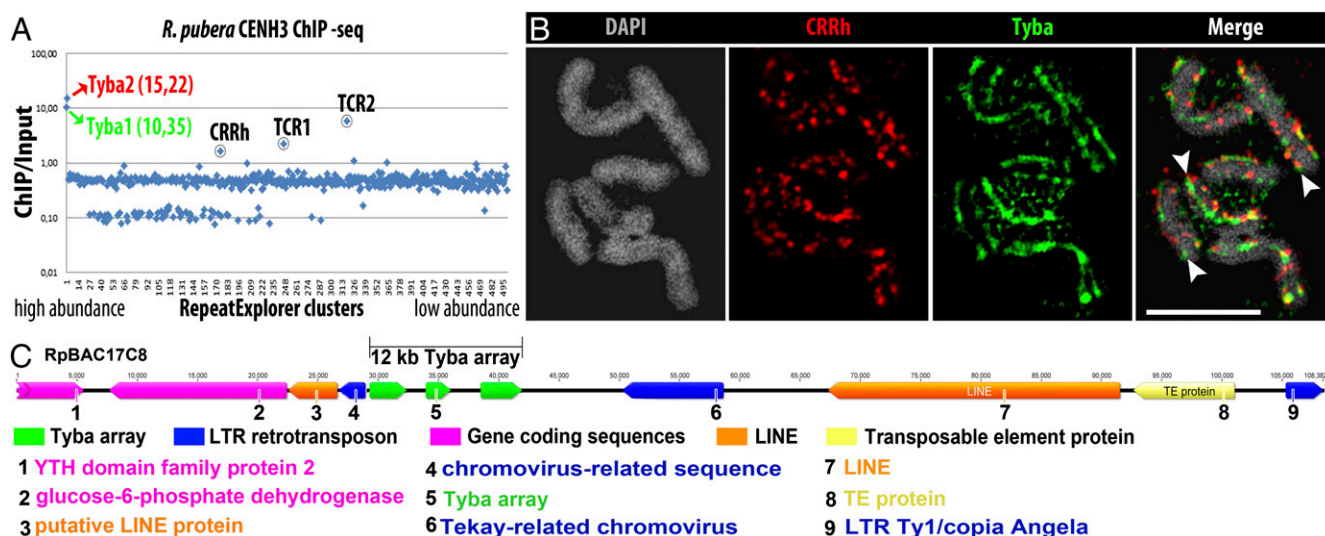


Fig. 3. Characterization of CENH3-interacting sequences. (A) CENH3 ChIP-seq reads mapped against the main RepeatExplorer clusters of the *R. pubera* genome. Colored circles and names indicate the main centromeric sequences in *R. pubera* CENH3 ChIP-seq. (B) SIM image showing FISH with CRRh (CL175) and Tyba1+2 on metaphase chromosomes. Arrowheads indicate the longitudinal centromere grooves. (Scale bar: 5 μ m.) (C) Annotation of an *R. pubera* BAC (RpBAC17C8) containing a centromere unit showing a Tyba array of ~12 kb divided into three subarrays inserted in the protein domains region of a chromovirus-related sequence. Additional transposable elements and single-copy coding sequences were found in close neighborhood.

DNA arrays in *Rhynchospora* compared to other species, the total amount of potential centromeric DNA per chromosome is among the largest reported for any species so far.

Conclusions

Although the mechanism behind the spreading of Tyba along the centromeric chromatin remains elusive, a preferential integration to CENH3-positive chromatin followed by positive selection might have occurred. On the other hand, the alternative option in which Tyba satellite repeat-rich regions may work as preferred sites for the deposition of centromeric nucleosomes and thus serve as potential kinetochore attachment sites in *R. pubera* cannot be excluded. Finally, the genome-wide distribution of centromeric repeat arrays interspersing the euchromatin observed in this species provides a previously unidentified variant of centromere organization. Thus, it is evident that different types of holocentromeres, namely centromeres with and without specific repetitive sequences and with or without CENH3/CENP-C, exist in different species. Further studies of species with holocentric chromosomes will broaden our mainly monocentric chromosome-biased knowledge about centromere organization and may help elucidate the centromere plasticity among eukaryotes.

Materials and Methods

Detailed materials and methods are described in *SI Appendix, Materials and Methods*. Briefly, high-copy repeats were identified by graph-based clustering (23) of genomic Illumina reads of *R. pubera*. A *Rhynchospora* CENH3-specific antibody was generated and used for indirect immunostaining, ChIP, ChIP-qPCR, and ChIP-seq. Centromeric repeats were characterized by FISH and sequence analysis.

ACKNOWLEDGMENTS. We thank Wayt Thomas for kindly identifying the plant material; Manuela Knauff for pulsed-field electrophoresis analysis; Karin Lipfert for art work; and Jan Vrána, Radka Tušková, and Eva Jahnová for flow sorting, BAC library construction, and filter preparation. The Brazilian Federal Agency for the Support and Evaluation of Graduate Education within the Ministry of Education of Brazil (CAPES) provided a Special Visiting Researcher Grant and project funding (to A.H.) and scholarships (to A.M. and T.R.). We thank the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) (AMD-0025-2.00-14) for visiting research grant (to A.M.). The Brazilian National Council of Technological and Scientific Development (CNPq) provided financial support for A.P.-H. The construction of the BAC library was funded by Ministry of Education, Youth and Sports of the Czech Republic National Program of Sustainability I Grant Award LO1204. The Czech Science Foundation and the Czech Academy of Sciences provided financial support for J.M. (GBP501/12/G090 and RVO:60077344).

- Plohl M, Meštrović N, Mravinac B (2014) Centromere identity from the DNA point of view. *Chromosoma* 123(4):313–325.
- Neumann P, et al. (2011) Plant centromeric retrotransposons: A structural and cytogenetic perspective. *Mob DNA* 2(4):1–16.
- Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: New insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* 82(2):261–282.
- Earnshaw WC, et al. (2013) Esperanto for histones: CENP-A, not CenH3, is the centromeric histone H3 variant. *Chromosome Res* 21(2):101–106.
- Heckmann S, Houben A (2013) Holokinetic centromeres. *Plant Centromere Biology*, eds Jiang J, Birchler JA (Wiley-Blackwell, Oxford, UK), pp 83–94.
- Guerra M, et al. (2010) Neocentrics and holokinetics (holocentrics): Chromosomes out of the centromeric rules. *Cytogenet Genome Res* 129(1–3):82–96.
- Moore LL, Morrison M, Roth MB (1999) HCP-1, a protein involved in chromosome segregation, is localized to the centromere of mitotic chromosomes in *Caenorhabditis elegans*. *J Cell Biol* 147(3):471–480.
- Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S (1999) A histone-H3-like protein in *C. elegans*. *Nature* 401(6753):547–548.
- Nagaki K, Kashiwara K, Murata M (2005) Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. *Plant Cell* 17(7):1886–1893.
- Heckmann S, et al. (2011) Holocentric chromosomes of *Luzula elegans* are characterized by a longitudinal centromere groove, chromosome bending, and a terminal nucleolus organizer region. *Cytogenet Genome Res* 134(3):220–228.
- Wanner G, Schroeder-Reiter E, Ma W, Houben A, Schubert V (June 6, 2015) The ultrastructure of mono- and holocentric plant centromeres: An immunological investigation by structured illumination microscopy and scanning electron microscopy. *Chromosoma*, 10.1007/s00412-015-0521-1.
- Cabral G, Marques A, Schubert V, Pedrosa-Harand A, Schlögelhofer P (2014) Chiasmatic and achiasmatic inverted meiosis of plants with holocentric chromosomes. *Nat Commun* 5:5070.
- Chmátal L, et al. (2014) Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr Biol* 24(19):2295–2300.
- Gassmann R, et al. (2012) An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. *Nature* 484(7395):534–537.
- Heckmann S, et al. (2013) The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. *Plant J* 73(4):555–565.
- Steiner FA, Henikoff S (2014) Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *eLife* 3:e02025.
- Subirana JA, Messegue X (2013) A satellite explosion in the genome of holocentric nematodes. *PLoS One* 8(4):e62221.
- Mello CC, Kramer JM, Stinchcomb D, Ambros V (1991) Efficient gene transfer in *C. elegans*: Extrachromosomal maintenance and integration of transforming sequences. *EMBO J* 10(12):3959–3970.
- Riddle DL, Blumenthal T, Meyer BJ, Priess JR (1997) Introduction to *C. elegans*. *C. elegans II*, eds Riddle DL, Blumenthal T, Meyer BJ, Priess JR (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 2nd Ed.
- Luceno M, Vanzela ALL, Guerra M (1998) Cytotaxonomic studies in Brazilian *Rhynchospora* (Cyperaceae), a genus exhibiting holocentric chromosomes. *Canadian Journal of Botany* 76(3):440–449.
- Vanzela ALL, Cuadrado A, Jouve N, Luceño M, Guerra M (1998) Multiple locations of the rDNA sites in holocentric chromosomes of *Rhynchospora* (Cyperaceae). *Chromosome Res* 6(5):345–349.
- Vanzela ALL, Guerra M, Luceno M (1996) *Rhynchospora tenuis* Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. *Cytobios* 88(355):219–228.
- Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11:378.
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29(6):792–793.
- Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409(1–2):72–82.
- Melters DP, et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* 14(1):R10.
- Kelly WG, et al. (2002) X-chromosome silencing in the germline of *C. elegans*. *Development* 129(2):479–492.
- Nagaki K, et al. (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* 163(3):1221–1225.
- Nagaki K, et al. (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36(2):138–145.
- Zhong CX, et al. (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14(11):2825–2836.
- Vafa O, Sullivan KF (1997) Chromatin containing CENP-A and alpha-satellite DNA is a major component of the inner kinetochore plate. *Curr Biol* 7(11):897–900.
- Hall LE, Mitchell SE, O'Neill RJ (2012) Pericentric and centromeric transcription: A perfect balance required. *Chromosome Res* 20(5):535–546.
- Iwata A, et al. (2013) Identification and characterization of functional centromeres of the common bean. *Plant J* 76(1):47–60.
- Dover GA (1986) Molecular drive in multigene families - How biological novelties arise, spread and are assimilated. *Trends Genet* 2(6):159–165.
- Mahtani MM, Willard HF (1998) Physical and genetic mapping of the human X chromosome centromere: Repression of recombination. *Genome Res* 8(2):100–110.
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191(4227):528–535.
- Talbert PB, Henikoff S (2010) Centromeres convert but don't cross. *PLoS Biol* 8(3):e1000326.
- Horvath JE, et al. (2005) Punctuated duplication seeding events during the evolution of human chromosome 2p11. *Genome Res* 15(7):914–927.
- Ma J, Jackson SA (2006) Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. *Genome Res* 16(2):251–259.
- Tek AL, Song J, Macas J, Jiang J (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. *Genetics* 170(3):1231–1238.
- Gorinsek B, Gubensek F, Kordis D (2004) Evolutionary genomics of chromoviruses in eukaryotes. *Mol Biol Evol* 21(5):781–798.
- Schrader F (1947) The role of the kinetochore in the chromosomal evolution of the heteroptera and homoptera. *Evolution* 1:134–142.
- Neumann P, et al. (2012) Stretching the rules: Monocentric chromosomes with multiple centromere domains. *PLoS Genet* 8(6):e1002777.
- C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* 282(5396):2012–2018.